

DOCUMENT RESUME

ED 466 498

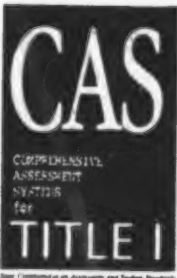
TM 034 237

AUTHOR Horm-Wingerd, Diane M.; Winter, Phoebe C.; Plofchan, Paula
TITLE Primary Level Assessment for IASA Title I: A Call for Discussion. Series on Standards and Assessments.
INSTITUTION Council of Chief State School Officers, Washington, DC.
SPONS AGENCY Office of Elementary and Secondary Education (ED), Washington, DC.
PUB DATE 2000-00-00
NOTE 47p.; Document produced with assistance from the State Collaborative on Assessment and Student Standards, Study Group on Early Childhood Assessment.
AVAILABLE FROM Council of Chief State School Officers, Attn: Publications, One Massachusetts Ave., N.W., Suite 700, Washington, DC 20001. Tel: 202-336-7016; Fax: 202-408-8072; Web site: <http://www.ccsso.org/publications>.
PUB TYPE Reports - Evaluative (142)
EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Accountability; Educational Assessment; *Elementary School Students; *Primary Education; *Student Evaluation; Test Construction; Test Interpretation; *Test Use
IDENTIFIERS *Improving Americas Schools Act 1994 Title I

ABSTRACT

The purpose of this paper is twofold: to review appropriate assessment techniques in prekindergarten through grade 3 settings and to serve as a catalyst for further discussion and work on the topic of developmentally appropriate accountability assessment. The discussion is based on the thesis that developmentally appropriate assessment and accountability assessment can be united. This thesis indicates that accountability assessment should be developmentally appropriate as defined in the National Association for the Education of Young Children and the National Association of Early Childhood Specialists in State Departments of Education (1991) joint position statement and should both yield information about individual children, schools, and districts and be linked with state standards. Despite the many challenges to development this type of assessment, a few local education agencies and state education agencies are trying to use appropriate assessments for accountability purposes. Drawing on their examples and research, recommendations are made to accomplish the goals of developmentally appropriate accountability assessment. These include: (1) forums to assess accountability assessment in early childhood programs; (2) demonstration projects to support the design, implementation, and testing of such programs; (3) funding to support research on appropriate methods of aggregating and interpreting assessment results; and (4) additional attention and resources devoted to professional development. An appendix discusses early childhood assessment. (Contains 46 references.) (SLD)

ED 466 498



Primary Level Assessment for IASA Title I A Call for Discussion

Diane M. Horm-Wingerd
Phoebe C. Winter
Paula Plofchan

with the
State Collaborative on Assessment and Student Standards
Comprehensive Assessment Systems for IASA Title I
Study Group on Early Childhood Assessment

PERMISSION TO REPRODUCE AND
DISSEMINATE THIS MATERIAL HAS
BEEN GRANTED BY
B. Buterbaugh
TO THE EDUCATIONAL RESOURCES
INFORMATION CENTER (ERIC)
1

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)
☒ This document has been reproduced as
received from the person or organization
originating it.
☐ Minor changes have been made to
improve reproduction quality.
• Points of view or opinions stated in this
document do not necessarily represent
official OERI position or policy.

TM034237



Council of Chief State School Officers
State Collaborative on Assessment and Student Standards

Primary Level Assessment for IASA Title I A Call for Discussion

Diane M. Horm-Wingerd
Phoebe C. Winter
Paula Plofchan

with the
State Collaborative on Assessment and Student Standards
Comprehensive Assessment Systems for IASA Title I
Study Group on Early Childhood Assessment

Council of Chief State School Officers
One Massachusetts Avenue, N.W.
Washington, DC 20001

© Copyright 2000 by the Council of Chief State School Officers. This document may not be reproduced without the permission of the Council of Chief State School Officers, the copyright holder.



The Council of Chief State School Officers (CCSSO) is a nationwide, nonprofit organization composed of the public officials who head departments of elementary and secondary education in states, the District of Columbia, the Department of Defense Education Activity, and five extra-state jurisdictions. CCSSO seeks its members' consensus on major education issues and expresses their views to civic and professional organizations, to federal agencies, to Congress, and to the public. Through its structure of standing committees and special task forces, the Council responds to a broad range of concerns about education and provides leadership on major education issues.

Because the Council represents each state's chief education administrator, it has access to the educational and governmental establishment in each state and to the national influence that accompanies this unique position. CCSSO forms coalitions with many other education organizations and is able to provide leadership for a variety of policy concerns that affect elementary and secondary education. Thus, CCSSO members are able to act cooperatively on matters vital to the education of America's young people.

The State Education Assessment Center is a permanent, central part of the Council. The Center was established through a resolution by the membership of CCSSO in 1984. This report is part of a series sponsored by the Assessment Center's State Collaborative on Assessment and Student Standards (SCASS), Comprehensive Assessment Systems for IASA Title I. The series addresses issues related to the standards and assessment provisions of Title I.

Preparation of this report was supported in part by the U.S. Department of Education, Office of Elementary and Secondary Education. The views and opinions expressed in this report are not necessarily those of the Council of Chief State School Officers, the U.S. Department of Education, or the State Collaborative on Assessment and Student Standards.

COUNCIL OF CHIEF STATE SCHOOL OFFICERS

Gordon M. Ambach
Executive Director

Wayne N. Martin
Director
State Education Assessment Center

John F. Olson
Director of Assessments
State Education Assessment Center

Phoebe C. Winter
Project Director, Comprehensive Assessment Systems for IASA Title I
State Collaborative on Assessment and Student Standards

**State Collaborative on Assessment and Student Standards
Comprehensive Assessment Systems for IASA Title I**

SERIES ON STANDARDS AND ASSESSMENTS

Opportunity and Change: An Overview of Title I of the Improving America's Schools Act of 1994. Frank Philip, 1995

Designing Coordinated Assessment Systems for Title I of the Improving America's Schools Act. Edward Roeber, 1995

Adequate Yearly Progress Provisions of Title I of the Improving America's Schools Act: Issues and Strategies. Dale Carlson, 1996

Implementing the Adequate Yearly Progress Provisions of Title I in the Improving America's Schools Act of 1994. Phoebe Winter, 1996

Using Existing Assessments for Measuring Student Achievement: Guidelines and State Resources. Linda Hansche, Tom Stubits, and Phoebe Winter, 1997

Analyzing, Disaggregating, Reporting, and Interpreting Students' Achievement Test Results: A Guide to Practice for Title I and Beyond. Richard Jaeger and Charlene Tucker, 1998

Handbook for the Development of Performance Standards: Meeting the Requirements of Title I. Linda Hansche, with contributions from Ronald Hambleton, Craig Mills, Richard Jaeger, and Doris Redfield, 1999

State Standards and State Assessment Systems: A Guide to Alignment. Paul M. La Marca, Doris Redfield, and Phoebe C. Winter, 2000

PREFACE

This report is the work of the State Collaborative on Assessment and Student Standards, Comprehensive Assessment Systems for IASA Title I (SCASS CAS) study group on early childhood assessment. It discusses appropriate assessment procedures for prekindergarten through grade 3, particularly as related to the program accountability requirements of Title I. Because of the dearth of appropriate operational models, the purpose of this paper is to showcase emerging approaches and to serve as a catalyst for further discussion and work in the area of early childhood accountability assessment. We hope that this paper will serve as a resource for state and local educational agency administrators as they search to identify or develop good primary-level assessments that can be used for Title I accountability. We also hope that this report helps explain to policymakers why these types of assessments are important and why further research and development efforts are needed.

SCASS CAS EARLY CHILDHOOD ASSESSMENT STUDY GROUP

Laurie Slobody (Co-chair)	Massachusetts Department of Education
Gail Taylor (Co-chair)	Vermont Department of Education
William Ashmore	Wisconsin Department of Public Instruction
Adrienne Bailey	Council of Great City Schools
Janet Carroll	United States Department of Education
David Frisbie	University of Iowa
Elaine Grainger	Vermont Department of Education
Janet Langlois	Louisiana Department of Education
Paula Plofchan	Council of Chief State School Officers
Grace Ross	United States Department of Education
Kathy St. Claire	Nevada Department of Education
Tom Stubits	Pennsylvania Department of Education
Jack Wills	Metropolitan Nashville Public Schools
Carole White	Delaware Department of Education
Phoebe Winter	Council of Chief State School Officers

The authors gratefully acknowledge the assistance of Hillary Michaels in providing information about and technical reviews of assessment instruments designed for use in the early grades.

EXECUTIVE SUMMARY

The purpose of this paper is twofold: to review appropriate assessment techniques in pre-K through grade 3 settings and to serve as a catalyst for further discussion and work on the topic of developmentally appropriate accountability assessment.

ASSESSMENT FOR TITLE I ACCOUNTABILITY

The 1994 reauthorization of the Elementary and Secondary Education Act made a significant number of changes to Title I, the compensatory education program that has served millions of American children and youth during its 30-plus years of existence. Some of the most significant changes occurred in the assessment and evaluation requirements of the law. Title I requires each state to develop or adopt a student assessment system to be administered annually to students in at least one grade in each of three grade ranges — grades 3 through 5, grades 6 through 9, and grades 10 through 12 — in at least mathematics and reading or language arts. All schools using Title I funds, even those with no grades in the specified ranges (e.g., a pre-K through grade 2 school), must provide information about their progress in assisting students to meet state performance standards.

A major reason that the law requires assessments to begin in grade 3 or above is the consensus of early childhood education experts that large-scale, group-administered, paper-and-pencil tests are inappropriate as the sole, or even primary, indicator of achievement for young children. However, Title I accountability requirements have several implications for assessment in the primary grades:

- the need to collect data on younger children to accomplish data-driven school reform in the context of high standards;
- the need to report progress of schools using Title I funds in grades K through 2 or 3;
- the recognition by Title I that large-scale, group-administered, paper-and-pencil tests are inappropriate for young children in preschool and primary programs; and
- the flexibility states and districts are allowed in choosing among options for student assessment to demonstrate Title I accountability.

APPROPRIATE ASSESSMENT OF YOUNG CHILDREN

This paper focuses on using appropriate assessments of young children for accountability purposes. The paper is based on the thesis that "developmentally appropriate assessment" and "accountability assessment" can be united. This thesis indicates that accountability assessment should be developmentally appropriate as defined in the National Association for the Education of Young Children (NAEYC) and the National Association of Early Childhood Specialists in State Departments of Education (1991) joint position statement; be designed to simultaneously yield meaningful information about individual children, schools, and districts; and be linked with state standards. Assessment of young children for the purpose of measuring school progress should

- be part of a broader system of continuous assessment of student learning;
- help teachers teach and students learn;
- support the curriculum and instructional goals of the school;
- be clearly tied to state content and performance standards;
- be appropriate for the ages and developmental levels of the children assessed; and
- be developed and used so that inappropriately narrow classroom instruction or other unintended negative consequences do not occur.

There are several challenges to using appropriate assessment for accountability purposes with young children:

- history of inappropriate over-reliance on norm-referenced tests;
- philosophical orientations and professional judgments;
- perceived conflict between "standards" and "developmentally appropriate practices;" and
- lack of training and experience.

Necessary conditions for addressing these challenges include

- increased understanding of standards and child-based results;
- valid interpretations of assessment results;
- increased understanding of appropriate assessment methods;
- increased understanding of the assessment process;

- the capability to aggregate results; and
- considerations for special populations.

EMERGING PRACTICES

Despite these challenges, a few local education agencies (LEAs) and state education agencies (SEAs) are attempting to use appropriate assessments for accountability purposes. For example, a district that has used developmentally appropriate assessment for both student-level and aggregate information is South Brunswick, New Jersey. In the 1997–98 school year, Vermont began using an individually administered reading assessment to determine the degree to which schools and districts are making progress in meeting standards in reading accuracy and comprehension. The Cambridge, Massachusetts, public school system has been developing and using an assessment program in early literacy since the early 1980s. Beginning with the 1998–99 school year, the Early Literacy Assessment, a performance-based assessment in reading and writing, replaced standardized testing in Cambridge in grades K through 2. The latter two programs are described in this paper to illustrate examples of emerging techniques for measuring young children for accountability purposes.

NEXT STEPS

State and local education agencies are encouraged to develop and support accountability policies and implementation strategies that incorporate developmentally appropriate assessments to provide information about young children's progress toward meeting state standards. Early childhood educators and educational measurement specialists must work together to derive satisfactory solutions to the need for developmentally appropriate assessments that can generate information to determine program effectiveness in supporting *all* young learners to achieve high standards. To accomplish the goal of developmentally appropriate accountability assessment in the early grades, it is recommended that:

1. forums be held to discuss accountability assessment in early childhood programs. Forums should be multidisciplinary and include professionals from the early childhood, assessment, educational measurement, and content areas including reading and math. We recommend these forums occur in the context of regular conferences of these professional groups. Additionally, we recommend that "summits" devoted to the exclusive discussion of developmentally appropriate accountability assessment be held. Summits should include policy makers, funders, and professionals representing a wide variety of disciplines related to early childhood development and learning.
2. demonstration projects be funded to support the design, implementation, and testing of approaches to developmentally appropriate accountability assessment. The requests for

proposals and contracts for these demonstration projects should contain stipulations that the models and results will be widely disseminated using a variety of formats (e.g., conferences, print material, Web-based technologies, etc.).

3. funding be directed to support research on appropriate methods of aggregating and interpreting results of developmentally appropriate assessment. Different techniques yielding different types of scores (e.g., referenced to developmental scales such as South Brunswick's Early Literacy Portfolio; referenced to standards such as the Vermont Developmental Reading Assessment) should be examined. Research should include investigations of the technical characteristics needed to use these techniques for aggregate decision-making and accountability purposes. For example, what degree of standardization in administration conditions is needed to obtain fair, valid, and reliable information about how a school is educating its students? What level of scorer consistency and score reliability is needed, and what are some ways to measure consistency appropriate to the nature of the instruments? This research can be conducted in conjunction with demonstration projects.
4. additional attention and resources be devoted to professional development on the topic of assessment of young children. Many of the challenges discussed in this paper relate directly to the lack of knowledge about assessment by most early childhood professionals and the lack of knowledge about early childhood pedagogy by most assessment specialists. Cross-fertilization of these two specializations is needed in preservice and inservice professional development programs. Incentives to achieve this cross-fertilization, such as special training grants, should be explored.

TABLE OF CONTENTS

PREFACE	I
EXECUTIVE SUMMARY	III
I. ASSESSMENT FOR TITLE I ACCOUNTABILITY	1
History	1
The 1994 Reauthorization of Title I	2
Implications for Pre-K through Grades 2 or 3	3
Title I Accountability Options for Pre-K through Grades 2 or 3	5
Developing and Using Appropriate Assessment of Young Children for Accountability	6
II. APPROPRIATE ASSESSMENT OF YOUNG CHILDREN	9
Early Childhood Assessment: General Definition and Purposes	9
Early Childhood Assessment: Concept of Developmentally Appropriate	9
Using Appropriate Assessment for Title I Accountability Purposes	10
The Challenges	10
Overcoming the Challenges	14
III. EMERGING PRACTICES	23
Vermont's Developmental Reading Assessment	23
Instrument Development	24
Administration and Scoring	24
Use in Instructional Planning and Monitoring	25
Use of Aggregated Results	25
Cambridge (MA) Public Schools' K-2 Early Literacy Assessment	25
Instrument Development	26
Administration and Scoring	26
Use in Instructional Planning and Monitoring	27
Use of Aggregated Results	27
IV. NEXT STEPS	29
Recommendations	29
APPENDIX A: EARLY CHILDHOOD ASSESSMENT	31
REFERENCES	35

I. ASSESSMENT FOR TITLE I ACCOUNTABILITY

HISTORY

In 1994, Congress reauthorized the Elementary and Secondary Education Act and made a significant number of changes to Title I, the compensatory education program that has served millions of American children and youth during its 30-plus years of existence. Some of the most significant changes occurred in the assessment and evaluation requirements of the law. These changes were prompted by the manner in which the assessment and evaluation components of the previous versions of the law had had unanticipated negative consequences on curriculum and assessment practices, as well as the federal government's desire to allow states and local districts to design standards and assessments that would help all students achieve at high levels. The reformers wanted to encourage states and school districts to build systems that would significantly raise student achievement for all students, including those served by Title I.

In the past, Title I (previously called Chapter I) evaluation and assessment practices helped "perpetuate inappropriate instructional methods and isolate Chapter I students from exciting and challenging experiences" (NAECS/SDE, 1991, p. 5). Although Chapter I regulations permitted flexible criteria for the selection of children to be served and did not require the use of standardized instruments for program evaluation until the beginning of second grade, most schools depended on large-scale, group-administered, norm-referenced tests for both selection of children and program evaluation. Many teachers and administrators observed that the over-reliance on group-administered, norm-referenced tests "constrained their ability to move toward more developmentally appropriate practices in the kindergarten and primary years" (NAECS/SDE, 1991, p. 3). This feeling was prevalent because of the widespread use of pullout programs, ability grouping, and subject fragmentation that were associated with the typical reliance on group-administered, norm-referenced tests for Chapter I purposes. Further, as noted by the National Association of Early Childhood Specialists in State Departments of Education, "the standardized tests drive not only the Chapter I services, but have a negative effect on the entire kindergarten and primary program in many schools across the nation" (NAECS/SDE, 1991, p. 3).

NAECS/SDE (1991) further noted that the typical approach to Chapter I assessment and instruction was in direct contrast to what was being advocated by early childhood experts and assessment reformers in the late 1980s and early 1990s. Child assessment that documented children's ongoing progress using multiple sources of information from actual learning situations and tasks obtained in meaningful contexts over time was advocated as the most appropriate form

of assessment. Regarding program evaluation, NAECS/SDE summarized the consensus of early childhood educators by stating that "Program evaluation should focus on the totality of children's development — social, emotional and physical — not just achievement which has historically been the Chapter I perspective" (p. 5). Research findings documenting the importance of critical inputs such as adult-child ratios and staff training in producing good outcomes for children (Harms and Clifford, 1993) led early childhood educators to advocate for the inclusion of such input variables in evaluations of programs designed for young children. These and other calls for change in assessment and evaluation procedures were influential in shaping the reauthorization of Chapter I.

THE 1994 REAUTHORIZATION OF TITLE I

The reauthorized Chapter I legislation, now Title I, requires each state to develop or adopt a student assessment system to be administered annually to students in at least one grade at the elementary, middle, and high school levels, in at least mathematics and reading or language arts. The measures used, and the standards they are based on, must be the same as those required of all students. States are given considerable flexibility to determine the makeup and format of content and performance standards, how they are developed, and how they are implemented. States had until the beginning of the 1997–98 school year to develop and implement their content and performance standards.

By the beginning of the 2000–2001 school year, states are expected to develop, pilot, and implement assessment systems based on their content and performance standards. These assessments will be the primary means for determining whether local education agencies (LEAs) and schools receiving Title I funds are serving students appropriately. Each state is required to establish criteria for determining whether LEAs and schools are making adequate yearly progress.

For Title I purposes, assessments must:

- be aligned with the state's content and performance standards;
- be administered at some time during grades 3 through 5, 6 through 9, and 10 through 12;
- include multiple measures of the state's content standards;
- include measures of higher order skills;
- be used for purposes for which the assessments are valid and reliable, and be consistent with relevant, nationally recognized professional and technical standards for those assessments;
- provide accommodations for students with diverse learning needs; and

- include students with limited English proficiency, in their primary language when practical, in a manner most likely to yield accurate and reliable information about them in subjects other than English.

Title I emphasizes the responsibility of schools and LEAs receiving Title I funds to educate all students to meet state standards. Schools and LEAs are accountable for the improvement they make each year in educating children served by Title I. If state assessments are not conducted in a Title I school, LEAs must use other measures or indicators to measure a school's progress (U.S. Department of Education, undated). This has particular implications for schools without any grade levels covered by the state assessment, typically primary schools.

States must develop procedures, or approve locally developed procedures, to determine whether schools and LEAs are making adequate yearly progress in enabling children served by Title I to meet the state's student performance standards. While states may allow locally developed procedures, states are responsible for ensuring that these procedures are appropriate and for defining adequate yearly progress.

Title I regulations [200.3(b)] require each state to define adequate yearly progress in a manner that:

- results in continuous and substantial yearly improvement of each school and LEA sufficient to achieve the goal of all children served under this subpart, particularly economically disadvantaged and limited English proficient children, meeting the State's proficient and advanced levels of performance;
- is sufficiently rigorous to achieve the goal within an appropriate timeframe; and
- links progress primarily to performance on the state's assessment system, while permitting progress to be established in part through the use of other measures, such as dropout, retention, and attendance rates.

In Title I targeted-assistance schools, measures of progress may be based on the performance of all students, only students who are currently served by Title I, or both students who are being served and students who have been served by Title I. In schools with Title I school-wide programs, the performance of all students must be used to determine the school's progress.

The requirements for measuring the progress of schools and LEAs allow states to collect information needed for Title I within a single system of accountability. States are encouraged to measure progress for Title I within the same accountability system they use to measure school and LEA progress in educating all students.

While states are developing their assessment and accountability programs, the yearly progress of schools and LEAs must be measured; however, the process may be different from that used after final assessments have been developed or adopted.

IMPLICATIONS FOR PRE-K THROUGH GRADES 2 OR 3

Several issues arise when considering Title I accountability policies and practices and their implications for pre-K through grades 2 or 3 assessment:

- The need to collect data on younger children to accomplish data-driven school reform in the context of high standards.

Recent research supports the importance of high-quality programs and instruction in the primary grades for forming the foundation for children's later school success. For example, the 1998 National Research Council report, "Preventing Reading Difficulties in Young Children," concludes "that quality classroom instruction in kindergarten and the primary grades is the single best weapon against reading failure" (p. 343). This research-based conclusion highlights the importance of program evaluation and improvement in the primary grades. Administrators require information prior to grades 3 or 4 for their planning and continuous improvement plans. There is a need for information that addresses the needs of all children, particularly the educational needs of historically underserved populations. Well-designed assessment programs should provide statistically sound, disaggregated data for whole school comprehensive planning. Without such data, the opportunity to strengthen programs for young children could be missed.

- The need to report progress of schools using Title I funds in Grades K through 2 or 3.

Although information from a state's school accountability system is not mandated prior to grade 3, there is a need to assess Title I programs designed for younger children. As noted in the U.S. Department of Education's *Guidance on Standards, Assessments and Accountability* (undated), over half the students served by Title I are in grades 3 or below. Thus, most of those served by Title I are enrolled in grades for which there may be no accountability information from a state's assessment system. If a state's assessment system begins in grade 4, for example, no state data will be available for primary grade students in targeted assistance schools or for students in schools with grade spans ending in grade 3 or below. However, each school receiving Title I funds must have some measure of annual progress for evaluation purposes, even if the school does not have any students who are included in the state's assessment system.

- The recognition by Title I that large-scale, group-administered, paper-and-pencil tests are inappropriate for young children in preschool and primary programs.

Title I does not require the use of the state's assessment system for school accountability until third grade or later. Because state accountability assessments often include large-scale, group-administered, paper-and-pencil tests, the position adopted by Title I is consistent with the consensus among early childhood educators that such tests are inappropriate for young children. The rapid, episodic, and individualistic development of young children, coupled with their lack of experience with the processes and goals of paper-and-pencil tests focused on discrete subject-oriented content, presents challenges

that are not as problematic in the assessment of older children and adults (Shepard, Kagan, and Wurtz, 1998). Guidelines for the appropriate assessment of young children have been developed (NAEYC and NAECS/SDE, 1991) and are discussed in Appendix A.

- The flexibility to choose among options for student assessment to demonstrate Title I accountability.

Title I does not require a specific approach to collecting student performance information to measure the progress of a school with grade levels below those covered by a state's mandated assessment system. Rather, Title I allows great flexibility for states and LEAs to document school progress. Specific options are described in the next section.

TITLE I ACCOUNTABILITY OPTIONS FOR PRE-K THROUGH GRADES 2 OR 3

Title I permits flexibility in selecting an approach to collect school progress information on the performance of children in the early grades who are enrolled in Title I programs. Four options are available for states to collect this information:

- tracking back;
- the SEA developing an appropriate assessment for young children;
- LEAs developing appropriate assessments for young children with SEA approval; and
- schools within an LEA developing appropriate assessments for young children with LEA and SEA approval.

With any of these options, it is possible and desirable to include additional indicators of student performance such as attendance, retention rates, etc. Information from more than one option can be combined to build an accountability system. The four options are discussed below.

1. Tracking back

If a state's assessment system begins in grade 4, it is possible for a K through 3 school to measure its progress by using the performance of fourth graders who attended that school. One method requires matching individual fourth-grade students' performance to each sending K through 3 school. Alternatively, the performance of fourth graders in a receiving school can be used to evaluate all feeder K through 3 schools that receive Title I funds, without matching students to a specific school. In targeted assistance schools with grade 4 students, performance of fourth-grade students currently served by Title I and performance of fourth-grade students previously served by Title I can be used to determine progress.

2. SEA develops appropriate assessment of young children

A state may opt to design an accountability system that includes children in the primary grades. For example, a state may elect to collect accountability assessment information from children enrolled in grade 2. If this option is selected, it is essential that the developers

design and implement an assessment system appropriate for young children in content, methods, and use.

3. LEAs develop appropriate assessment of young children with SEA approval
LEAs may develop or adopt system-wide assessments to be used below grade 4, with state approval. As with a statewide assessment system including young children, assessments developed and implemented by an LEA must be appropriate for young children in content, methods, and use.
4. Schools within an LEA develop appropriate assessment of young children with LEA and SEA approval
Individual schools may develop or adopt assessments to be used below grade 4, with LEA and SEA approval. Again, any assessments developed and implemented by an individual school must be appropriate for young children in content, methods, and use.

DEVELOPING AND USING APPROPRIATE ASSESSMENT OF YOUNG CHILDREN FOR ACCOUNTABILITY

This paper focuses on the last three options listed above: using appropriate assessments of young children for accountability purposes. Even in schools that include grades covered by the state's assessment system, additional information about the progress of students below the earliest assessed grade is critical to determine school progress effectively. While the first option of "tracking back" or using assessment data from higher grades provides some indication of school progress, more direct information about the achievement of students in earlier grades often is needed to provide a more complete picture about improvement. For schools with Title I programs focused on the early grades or without grade levels covered by a state's assessment system, the option of using assessment results from higher grades may not provide sufficiently specific information for measuring progress.

Options 2, 3, and 4, however, may appear to be in conflict with recent decisions of many states to abandon large-scale assessments of children below grade 3. This action has been based on the inappropriate use of large-scale, paper-and-pencil, group-administered testing of children below grade 3, consistent with the findings and recommendations of experts in the field of early learning. This paper is not advocating a return to these inappropriate practices. Rather, the intent of this paper is to illustrate the need to develop and use assessments appropriate for young children to inform school improvement. Schools, LEAs, SEAs, and the federal government need to know the effectiveness of programs designed to improve the learning and development of young children, including Title I students.

In the descriptions of options 2, 3, and 4 any assessments developed by schools, LEAs, or SEAs must be appropriate. Assessments designed to be used with young children should avoid methods such as group-administered, paper-and-pencil testing in favor of techniques that simulate classroom activities and enable children to provide a range of responses. In essence,

any state or locally designed measures must strive to meet the standards outlined by NAECS/SDE and NAEYC in their 1991 joint position statement entitled, "Guidelines for Appropriate Curriculum Content and Assessment in Programs Serving Children Ages 3 Through 8" (see Appendix A). This position statement and the topic of developmentally appropriate assessment with young children are discussed more fully in later sections of this paper.

While the focus of this paper is on Title I, the need for assessment information about early childhood programs to promote learning and the well-being of children is stressed by educators and policymakers. For example, CCSSO (1999), in its policy paper on early childhood and family education, underscores the need for collecting valid information about how programs serving young children are working.

Assessing young children to measure school progress should be part of a broader system of continuous assessment of student learning. Assessments should help teachers teach and students learn. Many of the assessment techniques designed for use with primary-level students have not been evaluated in the context of providing school-level results. States, LEAs, and schools should review a variety of strategies and determine which have the potential to provide the most useful and valid information, while supporting the curriculum and instructional goals of the school. Assessments must be clearly tied to state content and performance standards and be appropriate for the ages and developmental levels of the children assessed. Assessments must be developed and used so that they do not inappropriately narrow classroom instruction or produce other unintended negative consequences.

It is important to remember that the purpose of Title I is to ensure that *all* children, including those who have been disenfranchised in the past, master rigorous content. Any assessment tool used for any purpose, including that of measuring school progress, must be congruent with this fundamental goal. The challenge that school districts and states face is to appropriately assess students in the primary grades while producing results that can be used at the school level to measure school progress and inform instructional improvements.

II. APPROPRIATE ASSESSMENT OF YOUNG CHILDREN

EARLY CHILDHOOD ASSESSMENT: GENERAL DEFINITION AND PURPOSES

In 1991, the National Association for the Education of Young Children (NAEYC) and the National Association of Early Childhood Specialists in State Departments of Education (NAECS/SDE) released a joint position statement addressing appropriate curriculum content and assessment in programs serving children ages 3 through 8. In that document, assessment is defined as, "the process of observing, recording, and otherwise documenting the work children do and how they do it, as a basis for a variety of educational decisions that affect the child" (NAEYC and NAECS/SDE, 1991, p. 21). The document acknowledges that assessment is important in early childhood education because it serves four critical purposes or functions:

- to plan instruction for individuals and groups (classroom planning and individualization);
- to communicate effectively with parents;
- to identify children who may require specialized programs or intervention (screening and diagnostic assessment); and
- to provide information for program evaluation and accountability.

Each purpose or function, including a description of typical assessment methods, is discussed briefly in Appendix A.

EARLY CHILDHOOD ASSESSMENT: CONCEPT OF DEVELOPMENTALLY APPROPRIATE

In their joint position paper, NAEYC and NAECS/SDE (1991) emphasize that the concept of "developmentally appropriate" (Bredekamp, 1987; Bredekamp and Copple, 1997) applies not only to curriculum and instruction but also to the assessment of young children. A major portion that statement is devoted to guidelines for conducting developmentally appropriate assessment. These guidelines, reorganized by grouping similar items (i.e., guidelines on the content of assessment, on assessment methods or procedures, and on the outcomes of assessment), are reprinted in Appendix A. In general, appropriate assessment for young children relies on a variety of methods and permits several potential responses, is ongoing and summarizes performance over time, and consists of "hands-on" tasks that mirror classroom activities.

USING APPROPRIATE ASSESSMENT FOR TITLE I ACCOUNTABILITY PURPOSES

Historically, "developmentally appropriate" assessment has been used by classroom teachers to determine individual student needs and progress, inform instruction, and communicate with parents. The focus of this paper, however, is to encourage the development and use of appropriate early childhood assessment for "accountability" purposes. Although ample information is available on the classroom planning, communicating with parents, and screening/diagnostic purposes of assessment, few sources specifically address "developmentally appropriate" assessment for accountability purposes in early childhood settings. The few sources that do address this issue focus on the challenges inherent in such assessment. Given the daunting challenges and dearth of information, the purpose of this paper is to provide guidance and support to schools, LEAs, and SEAs as they seek to design systems to meet Title I accountability requirements.

This paper's thesis is that "developmentally appropriate assessment" and "accountability assessment" can be united. Accountability assessment should be developmentally appropriate as defined in the NAEYC and NAECS/SDE (1991) joint position statement; designed to simultaneously yield meaningful information about individual children, schools, and districts; and linked with state standards. Many would agree that this is an ideal conception of early childhood accountability assessment. Although the desirability of this type of assessment seems evident, much controversy and skepticism surround accountability assessment with young children (Kagan, Rosenkoetter, and Cohen, 1997).

The Challenges

It has been noted that the early childhood field has not sufficiently explored accountability assessment (Kagan et al., 1997). This lack of exploration of the desirability, feasibility, and procedures of accountability assessment is because many appear to be skeptical or wary of accountability assessment in early childhood settings. This wariness persists even though those in the field of early childhood education have acknowledged the importance of assessment to the design and implementation of developmentally appropriate programs for young children and their families. The wariness persists even though the field of early childhood education has achieved consensus regarding the appropriate content, methods, and use of assessments in early childhood settings (see Appendix A). The wariness persists even though the old problems associated with Chapter I practices have been acknowledged and addressed in the new law. Given the strides that have been made in early childhood assessment during the past decade, what is the cause of this continuing wariness?

Several factors underlie the skepticism and wariness that early childhood educators experience with assessment, especially accountability assessment. These factors include

1. history;
2. philosophical orientations and professional judgments;

3. perceived conflict between "standards" and "developmentally appropriate practices;" and
4. lack of training and experience.

Each of these factors is briefly discussed below.

1. History: Experiences and New Demands

The problems associated with the misuse and abuse of standardized tests with young children in the 1980s were well publicized. Many articles and conference presentations warned practitioners about the dangers of testing. Kamii (1990) published an influential book entitled *Achievement Testing in the Early Grades: The Games Grown-Ups Play* that chronicled common abuses and misuses of tests and the associated negative impacts on children, teachers, families, and school administrators. NAEYC published a position statement on tests and testing in the March 1988 issue of *Young Children*, a widely read journal for early childhood practitioners. Although these sources targeted real abuses and misuses of standardized testing, some over generalized and drew the incorrect conclusion that testing and assessment should be banned in early childhood settings. In fact, because of this widespread confusion, NAEYC issued a reexamination of the 1988 position statement in March 1989. In its reexamination, NAEYC noted that it "is not opposed to all standardized testing of young children and . . . is strongly in favor of ongoing assessment of children's development and learning as essential for appropriate curriculum planning and individualizing instruction" (p. 15). Even though NAEYC has been clear in its support of appropriate testing and assessment for a variety of purposes, including accountability, it is likely that continuing confusion and concern with test misuse make some hesitant about any type of assessment with young children. As noted by Schorr (1997), the field's experiences in the 1980s "have left the early childhood community so traumatized at the possibility of unwittingly promoting further inappropriate testing, that many oppose any attempt to assess school readiness by testing or observing individual children, even if the testing is done for the purpose of judging the community's provisions for preparing children for school entry, and not the abilities or capabilities of individual children" (p. 39). This attitude is not restricted to assessment of school readiness, but extends to all forms of early childhood assessment.

Contemporary pressures and challenges also underlie the uncertainty some early childhood educators experience with accountability assessment. Program evaluation and accountability assessment have only recently become more familiar concepts to most early childhood educators. With increased funding opportunities and emphasis on continuous improvement and documentation of outcomes, administrators of early childhood programs have just begun to think seriously about investing time and money in program evaluation and accountability assessment. As with any new enterprise, many are skeptical and unsure of appropriate procedures.

Given the dramatic growth in early childhood education and calls by Title I, the National Educational Goals Panel, and others to delineate goals and measure children's progress toward them, the resistance that early childhood educators have demonstrated toward accountability

assessment must be reexamined. There are signs that the field recognizes this. Forums on program evaluation and accountability assessment have been held recently (e.g., the Issues Forum on Child-Based Outcomes sponsored by the Kellogg Foundation, the Carnegie Corporation, and Quality 2000 and the resulting publication edited by Kagan et al., 1997), and leaders in the field have begun to discuss the importance of "child-based outcomes." There is growing recognition that, given the zeitgeist, accountability assessment of young children will occur with or without the input of the early childhood field. Thus, it is more productive for early childhood educators to participate in the discussion and actively shape the process. This paper encourages healthy collaboration between assessment and early childhood professionals to produce assessment systems that meet the mandates of Title I and support developmentally appropriate assessment and classroom practices for young children.

2. Philosophical Orientations and Professional Judgments

Professional philosophies and judgments affect views of accountability assessment. Even among those who are knowledgeable about assessment, various opinions exist about the appropriate use of the results of different types of assessments. There also is some disagreement regarding the wisdom of using classroom assessment results for accountability purposes. However, there is a growing consensus in the field that data gathered for one purpose may also be used for other purposes if appropriate planning and caution are displayed. For example, it is now accepted by some experts that classroom assessment data can be aggregated for accountability purposes (e.g., Roeber, 1996).

Professional views concerning children, their characteristics, and their development and learning also influence perceptions about the appropriateness of accountability assessment. As noted by Kagan et al. (1997), concerns are expressed about what is measured and how it is measured. Early childhood educators maintain that the content of assessment must include more than cognitive and pre-academic skills. Appropriate assessment must include developmental domains that are important to children's development and success but prove more difficult to measure. Examples include curiosity, persistence, motivation, and other socio-emotional constructs known to have an impact on development and learning. Many doubt that any single assessment can yield reliable or valid inferences about overall achievement because of the characteristics of young children, including their episodic learning, uneven growth, variability in behavior from setting to setting, and inexperience in "performing" in assessment situations.

Typical approaches to accountability assessment that include only child outcomes are in conflict with the basic philosophical orientation of many early childhood educators. The professional preparation of many early childhood professionals is based in a human ecological perspective that views the child as operating within spheres of influence that include the family, the school, the neighborhood, the wider community, the state, etc. In other words, the child's behavior and development are influenced by multiple factors in interactive and complex ways. Thus, it seems unnatural to collect assessment data exclusively on the child. This ecological philosophical orientation is supported by recent research demonstrating that important input variables such as teacher-child ratios and teacher training have impacts on child outcome variables (Burchinal, Peisner-Feinberg, Bryant, and Clifford, 1998; Harms and Clifford, 1993). In essence, many

early childhood professionals believe that accountability systems should include measures that go beyond child assessment data. As noted by Kagan et al. (1997), because of young children's developmental characteristics that increase the complexity of accurately assessing their capabilities, "ethical questions emerge regarding the legitimacy of basing child results (accountability data) only on what young children know and are able to demonstrate" (p. 8). Thus, it often is argued that results for young children must include more than what children know and can do. Many early childhood professionals contend that it is important to broaden the conception of results to include data that might be considered inputs of other age groups (Kagan et al., 1997). Examples include child health indicators, family income indicators, and other variables that affect what children know and can do.

3. Perceived Conflict Between "Standards" and "Developmentally Appropriate Practice"

Although standards have many proponents, some educators believe that academic content and performance standards are in conflict with the principles of "democratic education" (Noddings, 1997) and developmentally appropriate practices. For example, many believe that through standards, the same level of performance will be expected of all students (Noddings, 1997), the onus for learning (and the blame for not learning) will be placed exclusively upon the child (Darling-Hammond and Falk, 1997), or that failure and retention rates will be increased (Darling-Hammond and Falk, 1997). As noted by Darling-Hammond and Falk (1997), "Depending on how standards are shaped and used, either they could support more ambitious teaching and greater levels of success for all students, or they could serve to create higher rates of failure for those who are already least well-served by the education system" (p. 191). Thus, it is important to carefully craft standards and the associated assessments to support learning for all students. It is also critically important to use accountability information "to inform teaching decisions, to trigger special supports for student learning, and to evaluate school practices" (Darling-Hammond and Falk, p. 191).

It is important to recognize that accountability is about assessing student progress *and* continuously improving it (Darling-Hammond and Falk, p. 193). Thus, true accountability is tied to improvements in teaching and school practices that enhance the likelihood that students can meet standards (Darling-Hammond and Falk, 1997). Research demonstrates that developmentally appropriate practices enable young learners to meet high standards. As noted by Bredekamp and Copple (1997) in their NAEYC book entitled, *Developmentally Appropriate Practice in Early Childhood Programs*, teachers need to actively create intellectually engaging, responsive environments to promote individual learning and development. This is accomplished through a wide repertoire of teaching strategies and an integrated curriculum that has intellectual integrity and the opportunity to engage in experiential, hands-on learning. A common myth about developmentally appropriate practices is that the curriculum is "watered-down" and that the classroom lacks challenge and structure (Kostelnik, 1992). As noted by Kostelnik, a developmentally appropriate classroom is individualized to facilitate the optimal learning and development of each child.

4. Lack of Training and Experience

School personnel often lack the training and expertise to design and implement developmentally appropriate accountability assessment systems. For example, although it is acknowledged that reliance on group-administered, paper-and-pencil tests is inappropriate, administrators of early childhood programs often are unfamiliar with alternative assessment approaches. This is compounded by the apparent "ease" of group-administered, paper-and-pencil tests that can be easily purchased, machine scored, and offer computer-generated reports as convenient mechanisms to communicate assessment results. The challenge of designing and implementing more comprehensive and developmentally appropriate assessment approaches is often overwhelming to early childhood administrators who feel unprepared for the task (Thompson, 1990).

Early childhood educators also are often unfamiliar with assessment terms and approaches (Kagan et al., 1997). Preservice preparation programs often do not include in-depth training on all the types of assessments and their appropriate functions. Inservice programs are often "one-shot" seminars with little opportunity to implement new concepts with follow-up training. Thus, both new and seasoned teachers often are unfamiliar with the challenges of systematically including comprehensive assessment systems into their daily programs. Additionally, as noted by Kagan et al. (1997), limited consensus exists in the field of early childhood about what is meant by terms such as goals, benchmarks, indicators, results, assessment, and testing. So, like administrators, many teachers of young children feel uncomfortable with assessment, especially assessments that include high stakes for children and schools.

Misunderstandings are not restricted to the realm of assessment. Many misunderstand the concept of "developmentally appropriate practice," including its theoretical base, implementation, and relationship to standards. These misunderstandings have also affected the views of many early childhood educators regarding accountability assessment.

Overcoming the Challenges

The challenges described in the previous section must be addressed if "developmentally appropriate" assessments are to be used for early childhood accountability purposes. A variety of conditions must be put into place to support the design and use of developmentally appropriate early childhood assessments for Title I accountability purposes. The necessary conditions include:

1. increased understanding of standards and child-based results;
2. a focus on valid interpretations of assessment results;
3. increased understanding of appropriate assessment methods;
4. increased understanding of the assessment process;
5. the capability to aggregate results; and
6. considerations for special populations.

Each is discussed below in more detail.

1. Increased Understanding of Standards and Child-Based Results

Given the central role of standards in accountability systems, all educators should understand the impact of state standards on their work with young learners. Teachers must know and understand state content and performance standards and be proficient in using appropriate assessment techniques tied to those standards. Depending upon the starting point for a state's standards, developing early childhood assessments may require determining "benchmarks" for younger children that are based on the developmental continuum for a particular standard. This continuum would be "backmapped" from a state's formally assessed grade 3 or grade 4 standards. Emphasis should be placed on identifying where children are on the developmental continuum rather than stressing specific, predetermined grade expectations. In addition, there must be an understanding of early learning and a commitment to appropriate instruction and assessment on the part of the school and the LEA.

2. A Focus on Valid Interpretations of Assessment Results¹

A primary reason for using standards-based assessments is to measure student achievement against clear, comprehensive descriptions of what students know and can do along a continuum of learning. Assessment results should allow teachers and other users to understand where a child is in terms of knowledge, skills, and understanding in a particular content area, such as reading. The measurement techniques used must be able to support the inferences made based on the results. This is the core concept of validity: "the appropriateness, meaningfulness, and usefulness of specific inferences made from test scores" (AERA, APA, and NCME, 1985, p. 9). The misuse of standardized test scores noted earlier illustrates inappropriate interpretations and uses of test scores.

While, early childhood experts and assessment developers focused initially on the format of instruments used to measure the achievement of students in the primary grades, it is widely recognized that format per se is not an indicator of the degree of validity of an instrument (Baker, 1999). To validly assess children of any age, measurement techniques must be appropriate in terms of the cognitive demands placed on the student, be carefully tied to the outcomes being assessed, and measure knowledge and skills that are the results of learning and instruction. For young children, this means that a variety of techniques should be used, both for day-to-day instruction and at critical points for monitoring the effects of instruction. The level of technical quality needed will depend in large part on how the results will be used. For example, assessments used within the classroom as part of the teaching and learning process will not need to meet the same standards of technical quality as those used for reporting for program monitoring or accountability (Shepard et al., 1998).

Interpretations of assessment results must be based upon the constructs assessed and knowledge about how young children learn. As with all types of educational measurement, validity evidence should be collected to support the interpretations and inferences made from results of assessments used with primary grade children. For example, if results of an assessment are

¹ An extended discussion of the aspects of technical quality that should be considered before using any measurement technique is beyond the scope of this paper. The reader is referred to the *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 1999).

intended to describe what students know and can do in relation to specific mathematics content standards, evidence such as the degree of alignment of the assessment with the standards; the appropriateness of the measures for the students tested in term of cognitive complexity, language used, contextual information provided, etc.; and how well the results correspond to other measures of mathematical knowledge and skills² should be collected. This kind of information can help users determine how much faith they can put in various interpretations they wish to make based on the results. The stability of the results and their applicability to the breadth and variety of interpretations should be studied and carefully described in any reports of the results.

3. Increased Understanding of Appropriate Assessment Methods

To make valid inferences about student learning, especially with young children, it is important to use a variety of tools or techniques. Observations, interviews, and samples of work all can be useful in determining how well children are progressing. Assessment approaches should reflect good teaching and be integrated with classroom activities. Classroom-embedded assessment procedures provide immediate, familiar, and easily interpreted information to teachers. Such tasks do not necessarily mimic instructional activities, but they do reflect components of good early childhood learning environments. These tasks can provide data for measuring school progress, informing parents, and improving classroom practices.

An early childhood assessment system should include several methods of assessment. For example, language arts might be assessed through structured observations or oral reading, collections of samples of a student's writing, classroom-embedded tasks such as retelling a story, purposeful observations of a student's reading behavior (e.g., levels of books a child chooses), and interviews that assess oral proficiency. A distinguishing characteristic of these methods is that they are planned as assessment techniques, although they also may be used as instructional activities. Because they are assessments, they are closely tied to evaluative criteria based on content and performance standards and are designed and used to provide important information in a systematic, interrelated way. Methods should be systematically chosen to match content and performance standards and to ensure that a variety of methods are included in the overall assessment system.

Historically, accountability assessments have relied on paper-and-pencil, group-administered, multiple-choice tests. As discussed previously, these methods are inappropriate as the sole or even primary source of information about the performance of young children. The contemporary goal is to design accountability assessments that are developmentally appropriate. Developmentally appropriate methods for young children include techniques that "are based on actual observation and several samples of the child's work; include information gathered over time from a range of classroom experiences; indicate a child's broad progress in basic skills, conceptual understanding, problem solving, and reflective thinking, as well as motivation toward learning and attitudes toward school; are based on an understanding of developmental sequences and individual styles of learning; and are geared toward providing information that helps the

² There are a number of good sources on collecting evidence about the validity of inferences drawn from assessment results, including Messick (1989), and the aforementioned *Standards* (AERA, APA, and NCME, 1999).

teacher teach the individual child" (South Carolina Center for Excellence in the Assessment of Student Learning, 1993).

These two approaches — the traditional paper-and-pencil, group-administered, multiple-choice tests and developmentally appropriate assessments — can be thought of as two ends of a continuum. The following chart lists characteristics of assessment methods at the two ends of the continuum. The goal of this paper is to help readers move away from sole reliance on "traditional" methods and move toward incorporating developmentally appropriate assessment methods in their accountability assessment systems.

"Traditional" Approach	"Developmentally Appropriate" Approach
paper-and-pencil tasks	hands-on tasks that mirror classroom activities
multiple choice with one correct response	open-ended with several potential responses
administered once, captures performance at one given point in time	assessment is ongoing, summarizes performance over time to demonstrate growth
relies on one assessment method	incorporates several assessment methods
norms as the standard	developmental continua as the standard

A review of the appropriate methods to include in an early childhood assessment system is beyond the scope of this paper. Excellent sources can be consulted to investigate the characteristics, advantages, and disadvantages of assessment techniques. The reader is referred to sources such as McAfee and Leong (1997) and Bredekamp and Rosegrant (1992, 1995) for an overview of appropriate early childhood assessment techniques including observational methods of assessment such as anecdotal records, checklists, and rating scales; Grace and Shores (1992) for information on portfolio assessment, including the use of photographs, video recordings, audio recordings, or computer disks, in addition to written samples of student work; Meisels and Provence (1992) for information on screening; and Bailey and Wolery (1989), Meisels and Fenichel (1996), and Sattler (1992) for information on diagnostic assessment.

4. Increased Understanding of the Assessment Process

Great strides have been made in the past decade regarding our understanding of young children's learning (e.g., Wellman and Gellman, 1998) and appropriate ways to assess and document that learning. The challenge now is to disseminate that information to practitioners and to support their efforts to implement new assessment practices. This calls for systematic efforts to provide inservice professional development for current practitioners and to revise preservice preparation for aspiring teachers. Systematic professional development will increase practitioners' understanding of assessment, support their implementation of appropriate assessments, decrease the likelihood of assessments being misused or abused, and contribute to the attitude shift that is so important in helping early childhood educators recognize the value of accountability assessments.

Increased availability of model assessment systems demonstrating sound early childhood assessment practices and linked to state standards will also help remedy the issues identified

above. A major cause of the wariness of early childhood educators is the lack of familiarity with high quality, appropriate assessment approaches.

Teachers also need support to understand the developmental continua leading to the state standards for young children. This is especially true when state standards begin with third or fourth grade expectations with no attention given to benchmarks for younger children. Also, it is important for practitioners to understand and incorporate individual differences in development and to replace rigid grade-level expectations with the more flexible notion of mapping a child's progress along the developmental continuum. As noted previously, these notions are consistent with standards-based accountability and developmentally appropriate approaches to early childhood education.

(a) Administration

Most educational assessment programs routinely include training for teachers in administration procedures. Training becomes even more critical when the assessment techniques advocated in this paper are used as part of the accountability program. Fortunately, many of the skills needed to implement developmentally appropriate assessment techniques are the same as those teachers use in instruction. If the assessments are consistent with developmentally appropriate practice, and if teachers are familiar with the link between the assessments and the continua of learning goals (content standards), the foundation for professional development will already have been laid.

The degree and nature of the assessment's standardization and the specific techniques used (e.g., observation, portfolio of student work) will determine what types of professional development in administration procedures are needed. For example, a measure might require that teachers follow a written protocol verbatim; cover defined student knowledge or skills; and require that the teacher determine, for each child, the most appropriate point in the school year to administer the measure. Another measure might require teachers to collect evidence of student learning on specific content standards; provide a template outlining the types of admissible evidence; and develop standards-based instructional tasks and activities that allow for observation of the evidence needed. In the first example, the administration procedures and assessment content are standard for all students, but the timing might vary. (Some experts might argue that the variation in timing is a way of standardizing the assessment, in that the measure is administered at the appropriate time for each child.) In the second example, the template is for all across students, but the particular evidence used varies. (Again, some experts might argue that the variation standardizes the assessment, in that the tasks are appropriate to the teacher's particular instructional techniques, context, and classroom structure.)

In any case, teachers must have a clear understanding of assessment administration conditions. This can be especially important when the tasks used in the assessment are similar to tasks used in instruction. For example, if a portfolio assessment requires one piece of work to be an independent writing sample, it must be clear to the teacher (and student) that peer editing is not allowed for that piece of writing.

Professional development in the skills necessary to use the assessment techniques will depend, of course, on the particular technique(s) used. Teachers might need training in:

- using classroom observation instruments;
- taking running records;
- developing student portfolios;
- selecting and developing standards-based tasks; and
- determining the appropriate timing for an assessment.

Although many teachers may use these techniques regularly in instruction, they will need to be informed about how to use these same techniques in the context of a formal assessment program.

(b) Scoring

Some forms of developmentally appropriate assessment require teachers to make scoring judgments as part of administering the instrument, rather than simply collecting information that will be evaluated by outside scorers. In these cases, professional development in using standard scoring protocols for judging student work or observational information is critical. Even if assessments are scored by outside scorers, it is important to instruction and learning that teachers understand and can use the scoring techniques.

Consistency in applying scoring protocols is necessary when scores are used for accountability and other "high-stakes" purposes. A number of methods are available for checking on the consistency (or scorer reliability) of teacher scores and providing feedback, including the following techniques:

- other teachers score work from a sample of each teacher's students;
- two teachers observe the same student for a sample of students; and
- professional scorers score work from a sample of each teacher's students.

5. The Capability to Aggregate Results

If assessment results are to be used as part of a measure of a school's yearly progress in educating its students, results must be reported at a group level. While aggregated results are necessary for determining school progress, they also can provide information that is useful in making decisions about curriculum, program effectiveness, and instructional emphases. The uses of aggregated assessment results will determine the best ways to organize the data at the group level.

For the purpose of measuring school or district progress as required by Title I, assessment data typically are aggregated by subject area and grade level. For example, a school might report results for fourth graders in mathematics, language arts, and social studies. One way assessment

results can be aggregated in terms of predefined performances levels; for example, 20 percent of fourth graders scored at the "advanced" level on the state mathematics assessment, 42 percent scored at the "proficient" level, 31 percent scored at the "near proficient" level, and 7 percent scored at the "not proficient" level. When more than one assessment result is used to measure proficiency in a single subject area, a method of combining those results is used. At this time, research is under way to investigate various methods of combining results.

Aggregating results by grade level may not be appropriate for some pre-K through grade 3 assessment techniques. For example, at the school level, it may be more appropriate to report data on a developmental continuum across several grades, depending on the content standards and goals of the school. Performance levels might be reported for grades K through 2 combined using a scale such as the one used in the South Brunswick, New Jersey, Early Literacy Portfolio (Bridgeman, Chittendon, and Cline, 1995, p. 23 – 24):

1 Early Emergent

Displays an awareness of some conventions of writing, such as front/back of books, distinction between print and pictures. See[s] the construction of meaning from text as "magical" or exterior to the print. While the child may be interested in the contents of books, there is as yet little apparent attention to turning written marks into language. Is beginning to notice environmental print.

2 Advanced Emergent

Engages in pretend reading and writing. Uses reading-like ways that clearly approximate book language. Demonstrates a sense of a story being "read," using picture clues and recall of story line. May draw upon predictable language patterns in anticipating (and recalling) the story. Attempts to use letters in writing, sometimes in random or scribble fashion.

3 Early Beginning Reader

Attempts to "really read." Indicates beginning sense of one-to-one correspondence and the concept of word. Predicts actively in new material, using syntax and story line. Small sight vocabulary is becoming established. Evidence of initial awareness of beginning and ending sounds, especially in invented spelling.

4 Advanced Beginning Reader

Starts to draw on major cue systems; self-corrects or identifies words through use of letter-sound patterns, sense of story, or syntax. Reading may be laborious, especially with new material, requiring considerable effort and some support. Writing and spelling reveal awareness of letter patterns and conventions of writing such as capitalization and full stops.

5 Early Independent Reader

Handles familiar material on own, but still needs some support with unfamiliar material. Figures out words and self-corrects by drawing on a combination of letter-sound relationships, word structure, story line and syntax. Strategies of re-

reading or of guessing from larger chunks of texts are becoming well established. Has a large sight vocabulary. Conventions of writing are understood.

6 Advanced Independent Reader

Reads independently, using multiple strategies flexibly. Monitors and self-corrects for meaning. Can read and understand most material when the content is appropriate. Conventions of writing and spelling are — for the most part — under control.

Data need to be aggregated so that progress from year to year can be measured to determine how well schools are doing in improving their efforts to educate students. Data used to measure school and district performance must be tied to performance standards and content standards. This means that these standards and the reporting format should be closely aligned. For example, if it makes sense to have standards span several grade levels, it makes sense to measure school progress in increasing student proficiency across several grade levels.

For other purposes, such as improving curriculum, data would be analyzed differently. Data might be aggregated by grade level to see when students are reaching particular performance levels so that curricula can be geared appropriately. For instructional purposes, individual student results might be examined with reference to student work and teacher descriptions of other information related to student learning.

Title I also requires that assessment data be *disaggregated* by several defined categories: gender, race and ethnicity, English proficiency status, migrant status, disability status, and economic disadvantage status. Disaggregated results must be reported publicly, if it is statistically sound to do so. Jaeger and Tucker (1998) have written a report addressing appropriate ways to analyze, present, and use assessment results, with a focus on disaggregated results. While the report is written in non-technical terms for anyone in the public school system who deals with assessment results, it also includes statistical techniques for analyzing results.

6. Considerations for Special Populations

A basic tenet of standards-based school reform is that *all* children are included. To ensure equity and excellence, *all* students are expected to achieve the same challenging standards. To comply with Title I, states must use the same standards and assessments required of all students to determine progress of Title I programs. A range of accommodations often is needed to ensure that students with special needs can participate in assessment programs. Strategies have been developed and implemented to include students who are English language learners or who have disabilities (Butler and Stevens, 1997; Thurlow, Ysseldyke, and Olsen, 1999). Recent research suggests that these accommodation strategies do increase the inclusion of special populations. For example, Shepard (1998) reports that providing accommodations for students with limited English proficiency increases their participation in assessments and, consequently, enhances schools' ability to monitor their progress.

While schools and districts must provide opportunities for all children to be included in their assessments based on the best current information about the use of accommodations, consensus

exists that providing accommodations for students demonstrating limited English proficiency or disabilities is an area requiring further attention by researchers and assessment specialists. Shepard (1998) has noted that "the widespread use of accommodations raises questions about the variability of such accommodations and the validity of the assessment results, particularly since the use of accommodations seems to vary from school to school" (p. 5). Thus, the implications for schools, LEAs, and SEAs are that they should investigate the nature of accommodations and their associated outcomes as part of the development, use, data collection, and analysis of their assessment systems.

III.

EMERGING PRACTICES

Despite the challenges mentioned above, a few LEAs and SEAs are beginning to use appropriate assessments for accountability purposes. For example, South Brunswick, New Jersey,³ has used developmentally appropriate assessment for both student-level and aggregate information for several years. The Early Literacy Portfolio, used in grades K through 2, was developed by teachers in South Brunswick, with support from Educational Testing Service, to document students' emerging literacy skills. Teachers periodically collect student work samples to include in the portfolio, based on procedural guidelines. Portfolio contents are scored using a six-point, theory-based developmental scale (this scale is shown in an earlier section, "Aggregating Results"). Results are aggregated and used by the district to review how it is meeting its literacy-related goals and to allocate additional resources for students or groups of students.

This section describes techniques used by one state and one school district. These two examples are considered "emerging" because they are relatively recent and are different from approaches typically used for accountability purposes. Additionally, they are "models" in that they are two examples of "developmentally appropriate accountability assessments." However, they should be viewed as examples of what is possible, rather than as models to simply replicate. It is important for states, LEAs, and schools to review a variety of strategies and to design assessment systems that are tailored to their content standards and other unique features.

VERMONT'S DEVELOPMENTAL READING ASSESSMENT⁴

In the 1997-98 school year, Vermont began using an individually administered reading assessment to determine the degree to which schools and districts were making progress toward meeting standards in reading accuracy and comprehension. The VT-DRA is adapted from the Developmental Reading Assessment (published by Celebration Press), developed by Joetta Beaver of Upper Arlington, Ohio. The assessment uses individualized reading conferences and a series of books at increasing levels of difficulty so that teachers can determine the highest level of proficiency at which students read with accuracy and comprehension.

Teachers administer the VT-DRA to second-grade students in individual sessions lasting approximately 30 to 50 minutes. The student reads (silently, unless the student is at an early stage of reading) a short DRA book selected by the teacher. After reading, the student is asked to retell the story and respond to specific questions. To evaluate comprehension, teachers use a

³ Sources: Bridgeman, Chittenden, and Cline, 1995; Jones and Chittenden, 1995

⁴ Based on Biggam, 1998.

story-specific scoring guide to judge the adequacy of the retelling. To evaluate accuracy, the teacher takes and scores a record of oral reading, similar to a running record. If the student performs successfully on the proficiency level measured by the first book, the teacher has the student attempt a book at the next level (or "band"), until the teacher has determined the highest proficiency level at which the student reads with accuracy and comprehension.

Instrument Development

As part of the planning process, a task force of Vermont educators and other stakeholders selected the DRA based on criteria that included the assessment's connection to Vermont content standards in reading. After conducting feasibility studies, and with advice from the task force, the Vermont Department of Education worked with a team from the University of Vermont (supported by Brown University's Northeast and Islands Regional Education Laboratory) to revise the DRA to meet Vermont's large-scale assessment needs.

The Vermont Department of Education developed performance standards for the VT-DRA to measure student proficiency at five levels:

- Achieved the standard with honors (successful performance on books at a third-grade level or above);
- Achieved the standard (highest successful performance on books at a late second-grade level);
- Nearly achieved the standard (highest successful performance on books at a late first-grade to mid-second-grade level);
- Below the standard (highest successful performance on books at a primer to mid-first-grade level); and
- Little evidence of achievement (highest successful performance on books at an early first-grade level, or responded to familiar text).

The performance level earned is based on the student's highest performance on both reading accuracy and comprehension. During the 1998 school year, reading fluency was added as a requirement for proficiency at the highest (honors) level.

Administration and Scoring

Because the VT-DRA is part of Vermont's statewide assessment and accountability program, uniform administration and scoring are critical. Vermont requires six hours of initial training in administration and scoring, with attendance at updates and calibration sessions required annually for all who administer the assessment. Most schools use K through grade 3 teachers to administer the assessment. All administrations are audiotaped, and a sample of tapes is independently rescored by expert scorers. The expert scorers also evaluate the sample of audiotapes for consistency of administration, using a rubric designed to differentiate between valid and potentially invalidating administrations. Feedback is given to teachers whose

audiotapes are part of the sample. Vermont will conduct additional validity studies in conjunction with subsequent administrations of the VT-DRA.

Use in Instructional Planning and Monitoring

Because the VT-DRA is individually administered, it provides an opportunity for teachers to observe various aspects of students' developing reading: word identification strategies used or needed, fluency; proficiency in retelling stories and focusing on key ideas, and confidence during reading. Such information is useful both to refine instructional strategies for individual students and to guide possible shifts in curriculum emphasis. Schools are encouraged to send the information from the assessment to grade 3 teachers and to share this information with parents.

Use of Aggregated Results

Scores are aggregated at the school, district, and state levels for accountability purposes. Reports of the percentages of students who achieved each proficiency level are sent to each school and are available through the Vermont Department of Education's Web site. The aggregated results provide publicly available information about how schools are making progress toward having all students achieve early reading standards. Schools that demonstrate clear needs in early reading use the VT-DRA data and other indicators to set specific targets for improvement in student performance and plan strategies to effect change.

CAMBRIDGE (MA) PUBLIC SCHOOLS' K-2 EARLY LITERACY ASSESSMENT ⁵

The Cambridge, Massachusetts, Public School System has been developing and using an assessment program in early literacy since the early 1980s. Beginning with the 1998-99 school year, the Early Literacy Assessment (ELA) replaced district-wide standardized testing in grades K through 2. The ELA is a performance-based assessment in reading and writing administered formally in the fall and spring. The assessment consists of different components at different grade levels.

In kindergarten, letter identification, concepts about print, hearing sounds in words, and writing vocabulary are assessed using Marie Clay's Observation Survey (Clay, 1993). In grades 1 and 2, hearing sounds in words and writing vocabulary are similarly assessed. In addition, a modified version of the Developmental Reading Assessment (Engel, Hall, and Stuart, 1995) is used to assess oral reading and comprehension in grades 1 and 2. Formal writing assessments, through writing samples, are administered twice yearly and scored using a holistic scoring rubric. Cambridge also translates the assessments for students who are English language learners and has developed assessments in Spanish and Portuguese. The district is not yet using the assessments for formal accountability purposes but is investigating using the ELA and similar assessments in an accountability system.

⁵ Adapted from MacDonald and Hall, 1999, and Stuart, 1999

Instrument Development

From 1979 to 1988, Cambridge Public Schools focused on natural, developmental literacy acquisition. In 1983, the district began a five-year collaboration with Lesley College that included a longitudinal evaluation study of student progress in literacy. By providing teachers with opportunities to document student progress based on developmental assessment tasks in reading and writing, this evaluation laid the groundwork for the assessment development.

In 1988, the school district began to develop and implement portfolio assessment and longitudinal tracking of student progress from kindergarten through grade 12. Cambridge also initiated the Documentation and Assessment of Student Learning project, in which a group of primary teachers began to define the purposes and components of a system that gathered documentation for assessment from everyday classroom activities. As part of this project, a framework was developed that included portfolios, child-kept records, observations, inventories, and developmental assessment tasks in literacy and mathematics.

For the next 3 years, primary teachers used these assessment procedures in the classroom and provided additional refinements and adaptations. In 1993–94, the district completed the *Cambridge Handbook of Documentation and Assessment* (Engel et al., 1995). The outgrowth of this steady work on classroom assessment and documentation was the Early Literacy Assessment.

The reading assessment results are referenced to a reading continuum that spans grades K through 5. The continuum has 10 levels, from preconventional reader through experienced reader. The writing assessment results are referenced to the scoring rubric, which also spans grades K through 5 and has 10 levels.

Administration and Scoring

Attention is paid to ensuring that teachers can appropriately administer and score the assessments. From 1994 to 1997, teachers in each school participated in training sessions to learn the techniques of taking running records and analyzing the results. Teachers were also trained in scoring writing samples using the district's holistic writing rubric. Because each school in the district has support from an early childhood resource specialist, beginning in 1998–99, training in assessment became the responsibility of the school. Schools can request support from the district as needed.

Student writing samples are scored by school teams consisting of trained teachers. School teams do not score the assessments from students in their own schools. The scored samples are returned to teachers as soon as scoring is complete, accompanied by a class profile of scores. In addition, teachers are given an Excel database containing all reading and writing data for their classrooms.

Use in Instructional Planning and Monitoring

The ELA was designed to collect meaningful data that could be linked to classroom instruction. Teachers discuss ways the results can be used to plan individual instruction, and results are used in designing and revising curricula. The database of each student's reading and writing results allows teachers to look at individual student growth, class profiles, and trends in achievement. Schools are beginning to use the results to help develop individual school plans, reviewing the performance of students scoring below expectations in the fall and constructing instructional plans and interventions to assist those students.

Use of Aggregated Results

Results for each assessment component are aggregated by performance level for each classroom and school and for the district. The results are reported to the district's school committee in its annual Student Data Report. As noted, aggregate results are used for planning curriculum and school improvement. The district is beginning to develop an accountability system that can use the ELA and other curriculum-based assessment techniques, with the goal of moving accountability close to classroom practice so that the results are in context and are part of a coherent system for improving teaching and learning.

IV.

NEXT STEPS

Although the two examples described above are evidence that progress has been made in the effort to unite the concepts of "developmentally appropriate assessment" and "accountability assessment," much work remains. Currently, there is a paucity of operational models, and the literature contains very little information about appropriate accountability assessment of young children. However, it appears that the fields of early childhood education and educational measurement increasingly recognize that further work is needed. External pressures on both groups from legislators, funders, and citizens hopefully will encourage the two groups to bring about new, appropriate assessment approaches that will yield meaningful information to varied audiences. The danger is that increasing external pressure will prompt both groups to revert to old, inappropriate solutions: the implementation of paper-and-pencil, fact-driven, group-administered, norm-referenced tests for young children, and the insistence of early childhood educators that it is impossible to design and implement appropriate assessments for young children that can produce meaningful information for accountability purposes. The two groups must work together to derive satisfactory solutions that meet the need for developmentally appropriate assessments that can generate information useful in determining programs' effectiveness in supporting *all* young learners to achieve high standards.

RECOMMENDATIONS

The purpose of this paper is twofold: to review appropriate assessment techniques in pre-K through grade 3 settings and to serve as a catalyst for further discussion and work on the topic of developmentally appropriate accountability assessment. The paper does not provide definitive answers to the question of how best to collect information on the development and progress of children to gauge program success for Title I or other programs whose goals are to support young learners. Although there are some locally developed and commercially available instruments or assessment systems that are appropriate for use in the classroom, few have been developed with reference to content and performance standards or benchmarks, and few have been designed to yield aggregate scores. Thus, the following recommendations are made to prompt additional work in this area:

- Forums should be held to discuss accountability assessment in early childhood programs. Forums should be multidisciplinary and include professionals from the early childhood, assessment, educational measurement, and content areas including reading and math. We recommend these forums occur in the context of regular conferences of these professional groups. Additionally, we recommend that "summits" devoted to the exclusive discussion of developmentally appropriate

accountability assessment be held. Summits should include policy makers, funders, and professionals representing a wide variety of disciplines related to early childhood development and learning.

- Demonstration projects should be funded to support the design, implementation, and testing of approaches to developmentally appropriate accountability assessment. The requests for proposals and contracts for these demonstration projects should contain stipulations that the models and results will be widely disseminated using a variety of formats (e.g., conferences, print material, Web-based technologies, etc.).
- Funding should be directed to support research on appropriate methods of aggregating and interpreting results of developmentally appropriate assessment. Different techniques yielding different types of scores (e.g., referenced to developmental scales such as South Brunswick's Early Literacy Portfolio; referenced to standards such as the Vermont Developmental Reading Assessment) should be examined. Research should include investigations of the technical characteristics needed to use these techniques for aggregate decision-making and accountability purposes. For example, what degree of standardization in administration conditions is needed to obtain fair, valid, and reliable information about how a school is educating its students? What level of scorer consistency and score reliability is needed, and what are some ways to measure consistency appropriate to the nature of the instruments? This research can be conducted in conjunction with demonstration projects.
- Additional attention and resources should be devoted to professional development on the topic of assessment of young children. Many of the challenges discussed in this paper relate directly to the lack of knowledge about assessment by most early childhood professionals and the lack of knowledge about early childhood pedagogy by most assessment specialists. Cross-fertilization of these two specializations is needed in preservice and inservice professional development programs. Incentives to achieve this cross-fertilization, such as special training grants, should be explored.

APPENDIX A:

EARLY CHILDHOOD ASSESSMENT

EARLY CHILDHOOD ASSESSMENT: GENERAL DEFINITION AND PURPOSES

In 1991, the National Association for the Education of Young Children (NAEYC) and the National Association of Early Childhood Specialists in State Department of Education (NAECS/SDE) released a joint position statement addressing appropriate curriculum content and assessment in programs serving children ages 3 through 8. In that joint document, assessment was defined as, "the process of observing, recording, and otherwise documenting decisions that affect what children do and how they do it, as a basis for a variety of educational decision that affect the children (NAEYC & NAECS/SDE, 1991, p. 21). These associations acknowledge that assessment is very important in early childhood education because it serves four purposes or functions:

1. to plan instruction for individuals and groups (classroom planning and individualization);
2. to effectively communicate with parents;
3. to identify children who may require specialized programs or interventions (screening and diagnostic assessment);
4. to provide information for program evaluation and accountability.

Each purpose of function, including a description of typical methods, is briefly discussed below.

Classroom planning: In classroom planning, assessment information is used by teachers to design activities to meet the needs of individual children and to guide overall classroom planning. Such assessments typically occur on an ongoing basis and involve techniques such as observation, checklists, anecdotal records, or portfolios. An example of such assessments include teacher observation of a child's small motor development to plan activities that foster the child's development (Kagan, Rosenkoetter, & Cohen, 1997).

Communicating with parents: For this purpose, assessment information is collected by teachers in order to give families a comprehensive picture of their children's ongoing development. Assessment techniques used by the teacher often include observation, checklist, logs, or portfolios with the purpose of demonstrating growth over time. An example of this type of assessment would be a teacher reviewing writing samples drawn from a child's portfolio to highlight developmental changes over time as part of a parent-teacher conference (McAfee & Leong, 1997).

Screening and diagnostic evaluation: In this purpose, assessment information is collected to identify children with specific characteristics, describe children's current levels of functioning,

and determine eligibility for intervention services. Screening involves assessing a large number of children quickly to identify those who may have special needs. Diagnostic assessment is more thorough and comprehensive assessment to document the child's characteristics and to determine if service eligibility criteria are met. Observation, parent interviews, and tests are commonly used assessment approaches (Kagan et al., 1997; Meisels & Provence, 1992).

Program evaluation: For this purpose, assessment is conducted to determine the impact of a specific program or a particular intervention. The results are usually used to guide future programming and funding decisions. Typically, the performance of a group of children is of interest and this type of assessment is often conducted by an evaluator or researcher rather than classroom personnel. Examples of this type of assessment includes the evaluation of the Parents as Teachers program (Kagan et al., 1997).

Accountability: Accountability assessment is conducted to inform the public about the collective status of children. For this purpose, groups of children are the unit of study, the performance of children in classrooms, schools, districts, communities, states, and nations is the focus. Assessment must be relatively time-efficient and the results must be comparable and capable of aggregation. Results from this type of assessment tend to be broadly disseminated and used for decision-making. Thus, assessment for accountability purposes is "high stakes" at the school level and higher (Kagan et al., 1997).

CONCEPT OF DEVELOPMENTALLY APPROPRIATE

In their joint position paper, NAEYC and NAECS/SDE (1991) emphasized that the concept of "developmentally appropriate" (Bredekamp, 1987; Bredekamp & Copple, 1997) applies not only to curriculum and instruction but also to the assessment of young children. A major portion of their 1991 joint position statement was devoted to listing guidelines for conducting developmentally appropriate assessment. These guidelines, reorganized by grouping similar items (i.e., guidelines on the content of assessment, guidelines on assessment methods or procedures, and guidelines on the outcomes of assessment), are reprinted in the next section entitled, "Guidelines for Assessment."

In conjunction with other influences (e.g., federal mandates such as P.L. 99-457), the 1991 NAEYC and NAECS/SDE guidelines served as a catalyst for increased attention and concern with the appropriate assessment of young children. In fact, information concerning assessment for classroom planning, communicating with parents, and screening and diagnostic purposes became readily available in the mid to late 1990s. Although a comprehensive listing is beyond the focus of this document, several excellent sources are currently available for readers interested in assessment in early childhood education. For example, see Bredekamp and Rosegrant (1992, 1995) and McAfee and Leong (1997) for additional information related to assessment for classroom planning and communicating with parents; see Meisels and Provence (1992) for information on screening and Meisels and Fenichel (1996), Bailey and Wolery (1989), and Sattler (1992) for information on diagnostic assessment with young children.

GUIDELINES FOR ASSESSMENT: EXCERPTED AND REORGANIZED FROM NAEYC AND NAECS/SDE, 1991

NAEYC and NAECS/SDE (1991) in their joint position statement entitled, "Guidelines for Appropriate Curriculum Content and Assessment in Programs Serving Children Ages 3 through 8," have listed guidelines for assessment when planning instruction and communicating with parents.

Guidelines Concerning the Content of Assessment:

- Curriculum and assessment are integrated throughout the program; assessment is congruent with and relevant to the goals, objectives, and content of the program.
- Children's development and learning in all the domains—physical, social, emotional, and cognitive—and their dispositions and feelings are assessed.
- Assessment relies on demonstrated performance, during real, not contrived activities, for example, real reading and writing activities rather than only skills testing.
- Assessment recognizes individual diversity of learners and allows for differences in styles and rates of learning. Assessment takes into consideration children's ability in English, their stage of language acquisition, and whether they have been given the time and opportunity to develop proficiency in their native language as well as in English.
- Assessment demonstrates children's overall strengths and progress, what children can do, not just their wrong answers or what they cannot do or do not know.
- Assessment addresses what children can do independently and what they can demonstrate with assistance, since the latter shows the direction of their growth.

Guidelines Concerning Methods or Procedures:

- Assessment involves regular and periodic observation of the child in a wide variety of circumstances that are representative of the child's behavior in the program over time.
- Assessment relies primarily on procedures that reflect the ongoing life of the classroom and typical activities of the children. Assessment avoids approaches that place children in artificial situations, impede the usual learning and developmental experiences in the classroom, or divert children from their natural learning processes.
- Assessment utilizes an array of tools and a variety of processes including but not limited to collections of representative work by children (artwork, stories they write, tape recordings of their reading), records of systematic observations and interviews of children's progress as individuals and as groups.
- Information about each child's growth, development, and learning is systematically collected and recorded at regular intervals. Information such as samples of children's

work, descriptions of their performances, and anecdotal records are used for planning instruction and communicating with parents.

- Assessment is an essential component of the teacher's role. Since teachers can make maximal use of assessment results, the teacher is the primary assessor.
- Assessment encourages children to participate in self-evaluation.
- Assessment is a collaborative process involving children and teachers, teachers and parents, school and community. Information from parents about each child's experiences at home is used in planning instruction and evaluating children's learning. Information obtained from assessment is shared with parents in language they can understand.

Guidelines Concerning the Outcomes of Assessment:

- Assessment results in benefits to the child such as needed adjustment in the curriculum or more individualized instruction and improvements in the program.
- Assessment provides teachers with useful information to successfully fulfill their responsibilities: to support children's learning and development, to plan for individuals and groups, and to communicate with parents.
- Assessment supports children's development and learning. It does not threaten children's psychological safety or feelings of self-esteem.
- Assessment supports parents' relationships with their children and does not undermine parents' confidence in their children's or their own ability, nor does it devalue the language and culture of the family.
- A regular process exists for periodic information sharing between teachers and parents about children's growth and development and performance. The method or reporting to parents does not rely on letter or numerical grades, but rather provides more meaningful, descriptive information in narrative form.

REFERENCES

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1985). *Standards for educational and psychological testing*. Washington, DC: American Psychological Association.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Psychological Association.

Bailey, D. B., & Wolery, M. (1989). *Assessing infants and preschoolers with handicaps*. Columbus, OH: Merrill.

Baker, E. (1999, June). Discussant remarks. In D. Horm-Wingerd (moderator), *Assessment and accountability in the early grades*. Symposium presented at the 29th Annual National Conference on Large-Scale Assessment, Snowbird, UT.

Biggam, S. C. (1998). Personal communication.

Bredekamp, S. (Ed.). (1987). *Developmentally appropriate practice in early childhood programs serving children from birth through age 8*. Washington, DC: NAEYC.

Bredekamp, S., & Copple, C. (Eds.). (1997). *Developmentally appropriate practice in early childhood programs*. Washington, DC: NAEYC.

Bredekamp, S., & Rosegrant, T. (Eds.). (1992). *Reaching potentials, Volume 1: Appropriate curriculum and assessment for young children*. Washington, DC: NAEYC.

Bredekamp, S., & Rosegrant, T. (Eds.). (1995). *Reaching potentials, Volume 2: Transforming early childhood curriculum and assessment*. Washington, DC: NAEYC.

Bridgeman, B., Chittenden, E., & Cline, F. (1995). *Characteristics of a portfolio scale for rating early literacy*. Princeton, NJ: Educational Testing Service.

Burchinal, P., Peisner-Feinberg, E., Bryant, D., & Clifford, D. (1998). *Children's social and cognitive development and child care quality: Testing for differential associations related to poverty, gender, or ethnicity*. Chapel Hill, NC: The National Center for Early Development and Learning.

Butler, F. A., & Stevens, R. (1997). *Accommodation strategies for English language learners on large-scale assessments: Student characteristics and other considerations*. CSE Technical Report 448. Los Angeles: CRESST.

Clay, M. M. (1993). *An observation survey of early literacy achievement*. Portsmouth, NH: Heinemann.

Council of Chief State School Officers. (1999). *Early childhood and family education: New realities, new opportunities*. A council policy statement. Washington, DC: Author.

Darling-Hammond, L., & Falk, B. (1997). Using standards and assessments to support student learning. *Phi Delta Kappan*, 79(3), 190-199.

Engel, B., Hall, L., & Stuart, L. (Eds.). (1995). *The Cambridge handbook of documentation and assessment: Child portfolios and teacher records in the primary grades*. Grand Forks, ND: North Dakota Study Group on Evaluation, University of North Dakota.

Grace, K., & Shores, E. (1992). *The portfolio and its use*. Little Rock, AR: Southern Association on Children Under Six (SACUS).

Harms, T., & Clifford, R. M. (1993). Studying educational settings. In B. Spodek (Ed.), *Handbook of research on the education of young children* (p. 477-492). New York: Macmillan.

Jaeger, R. M., & Tucker, C. (1998). *Analyzing, disaggregating, reporting and interpreting students' achievement test results: A guide to practice for Title I and beyond*. Washington, DC: Council of Chief State School Officers.

Jones, J., & Chittenden E. (1995). *Teachers' perceptions of rating an early literacy portfolio*. Princeton, NJ: Educational Testing Service.

Kagan, S. L., Rosenkoetter, S., & Cohen, N. (1997). *Considering child-based results for young children: Definitions, desirability, feasibility, and next steps*. New Haven, CT: Yale Bush Center in Child Development and Social Policy.

Kamii, C. (Ed.). (1990). *Achievement testing in the early grades: The games grown-ups play*. Washington, DC: NAEYC.

Kostelnik, M. (1992). MYTHS associated with developmentally appropriate programs. *Young Children*, 47(4), 17-23.

Koretz, D. (1997). *The assessment of students with disabilities in Kentucky*. CSE Technical Report 431. Los Angeles: CRESST.

MacDonald, M., & Hall, L. (1999). *K-2 early literacy assessment in Cambridge Public Schools*. Report to the Massachusetts Department of Education. Cambridge, MA: The Cambridge School Department.

McAfee, O., & Leong, D. (1997). *Assessing and guiding young children's development and learning*. Boston: Allyn & Bacon.

Meisels, S. J., & Fenichel, E. (Eds.). (1996). *New visions for the developmental assessment of infants and young children*. Washington, DC: Zero to Three, the National Center for Infants, Toddlers, and Families.

Meisels, S. J., & Provence, S. (1992). *Screening and assessment: Guidelines for identifying young disabled and developmentally vulnerable children and their families*. Washington, DC: Zero to Three, the National Center for Infants, Toddlers, and Families.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational Measurement* (Third Edition). New York: American Council on Education and MacMillan Publishing Company.

National Association of Early Childhood Specialists in State Departments of Education. (1991). Chapter I services and early childhood education: Problem or promise. A position statement prepared by the National Association of Early Childhood Specialists in State Department of Education.

National Association for the Education of Young Children. (1988). Position statement on standardized testing of young children 3 through 8 years of age. *Young Children*, 43(3), 42-47.

National Association for the Education of Young Children. (1989). A reexamination of NAEYC's position on standardized testing. *Young Children*, 44(3), 15.

National Association for the Education of Young Children and National Association of Early Childhood Specialists in State Departments of Education. (1991). Guidelines for appropriate curriculum content and assessment in programs serving children ages 3 through 8. *Young Children*, 46(3), 21-38.

National Research Council. (1998). *Preventing reading difficulties in young children*. Washington, DC: National Academy Press.

Noddings, N. (1997). Thinking about standards. *Phi Delta Kappan*, 79(3), 184-189.

Roeber, E. (1996). *Designing coordinated assessment systems for Title I of the Improving America's Schools Act of 1994*. Washington, DC: Council of Chief State School Officers.

Sattler, J. (1992). *Assessment of children*. Boston: Allyn & Bacon.

Schorr, L. (1997). Judging interventions by their results. In S. L. Kagan, S. Rosenkeotter & N. Cohen (Eds.), *Considering child-based results for young children: Definitions, desirability, feasibility, and next steps*. New Haven, CT: Yale Bush Center in Child Development and Social Policy.

Shepard, L. (1998, Fall). 1998 CRESST Conference Focused on Urgent Issues. *The CRESST Line: Newsletter of the National Center for Research on Evaluation, Standards, and Student Testing*, p. 4-5.

Shepard, L., Kagan, S. L., & Wurtz, E. (1998). *Principles and recommendations for early childhood assessments*. Washington, DC: National Education Goals Panel.

South Carolina Center for Excellence in the Assessment of Student Learning. (1993). Contact CEASL, College of Education, University of South Carolina, Columbia, SC 29208.

Stuart, L. (1999). Personal communication.

Thompson, E. W. (1990). The dilemma for superintendents. In C. Kamii (Ed.), *Achievement testing in the early grades: The games grown-ups play*. Washington, DC: NAEYC.

Thurlow, M., Ysseldyke, J., & Olsen, K. (1999). *Self-study guide for the development of statewide assessments that include students with disabilities*: Minneapolis, MN: National Center on Educational Outcomes.

U.S. Department of Education (undated). *Guidance on standards, assessments, and accountability*. Washington, DC: Author.

Wellman, H. M., & Gellman, S. A. (1998). Knowledge acquisition in foundational domains. In W. Damon (Editor-in-chief) and D. Kuhn & S. Siegler (Vol. Eds.), *Handbook of child psychology: Vol. 2. Cognition, perception, and language* (5th ed., p. 523-573). New York: Wiley.



U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

TM034237



NOTICE

Reproduction Basis



This document is covered by a signed "Reproduction Release (Blanket)" form (on file within the ERIC system), encompassing all or classes of documents from its source organization and, therefore, does not require a "Specific Document" Release form.



This document is Federally-funded, or carries its own permission to reproduce, or is otherwise in the public domain and, therefore, may be reproduced by ERIC without a signed Reproduction Release form (either "Specific Document" or "Blanket").